

### Short-guide to the AP-MS bioinformatics pipeline

1. Filtering out the MS output for protein groups that could not be uniquely matched to a set of peptide spectra in the search database
2. Computationally removing carry-over proteins observed between IPs that were run in consecutive order
3. Clustering of total protein spectral count (or # of unique peptides) correlation between all IPs to identify
  - a. Clusters of replicates (good)
  - b. Dispersed individual IPs that cluster more with the negative controls than with their replicates (bad)
4. Optional: Remove 'bad' IPs from the dataset
5. Scoring (See below)
  - a. MIST with virus-host parameters
  - b. MIST with PCA derived parameters
  - c. COMPPASS
  - d. (Optional: SAINT)
6. Optional: Adding additional background from AP-MS database to improve detection of commonly observed unspecific binders and rerun 5.
7. Optional: Searching for enriched complexes or pathways for a given bait

### Short-guides to 'scored' table format

- BAIT: custom bait name
- PREY: uniprot\_ac code for identified prey protein
- MIST\_hiv: MIST score with 'HIV' weights for this bait-prey pair (higher = more significant)
- MIST\_hiv\_negative: MIST score with 'HIV' weights for this prey against negative controls (specific interactions should have a low negative score)
- MIST\_self: MIST score with custom weights derived by PCA, the higher the more significant (higher = more significant)
- MIST\_self\_negative: MIST score with custom weights for this prey against negative controls (specific interactions should have a low negative score)
- MIST\_R: MIST reproducibility feature
- MIST\_A: MIST abundance feature
- MIST\_S: MIST specificity feature
- TSC\_AVG: Average Total Spectral Count (or # unique peptides) over replicates
- COMPPASS\_Z: Z-score for this bait-prey pair's TSC\_AVG with respect to observations of other baits interacting with this prey
- COMPPASS\_S: Empirical COMPPASS Specificity score for this bait-prey pair (higher = more significant)
- COMPPASS\_D: Empirical COMPPASS Specificity & Reproducibility score for this bait-prey pair (higher = more significant)
- COMPPASS\_WD: Empirical COMPPASS Specificity & Reproducibility score with background correction for this bait-prey pair (higher = more significant)
- COMPPASS\_WD\_negative: Empirical COMPPASS Specificity & Reproducibility score with background correction for this pair against the negative controls (specific interactions should have a low negative score)
- SAINT\_AVG\_P: Average SAINT score over all replicates for this bait-prey pair
- SAINT\_MAX\_P: Maximal SAINT score over all replicates for this bait-prey pair
- IPs: commas separated list of all IP identifiers where this prey was identified in
- Remaining columns: Uniprot descriptions for preys

#### Note on thresholds:

1. We suggest a MIST threshold of  $> 0.7$  for significant interactions
2. COMPPASS score are not intuitive to interpret but we propose ranking any of the reported scores descending and setting the 95%-ile as a score threshold
3. SAINT scores correspond to the probability of an interaction being true. The authors suggest a threshold of  $> 0.9$

#### Short-guides to score calculations

##### MIST

The MIST algorithm first computes three features for every bait-prey interaction given the whole input set of observed interactions: *abundance*, *reproducibility* and *specificity*. The MIST total reported score is a weighted sum of these three features. The weights were determined by Principal Component Analysis (PCA) to maximize the feature space variance in one dimension.

We currently report a MIST score with the optimal parameters for the HIV-host interaction networks. We postulate that these parameters are suitable for most sparsely interconnected bait-prey datasets (eg. virus-host interaction networks).

We also report the MIST score with custom computed weights by performing PCA on the input set of interactions. For a more detailed description of the MIST algorithm we refer to the online published supplementary material of the '[Global landscape of HIV-human protein complexes paper](#)'

##### COMPPASS

An excellent description of the COMPPASS score can be found online on the [Harper Lab website](#)

##### SAINT

All the information regarding the SAINT scoring algorithm can be found online on the [SAINT website](#)